...el Research and Development Center

(2)

Development of Performance Measures
for the Navy Radioman

AD-A198 984

DTIC
SELECTE
S SEP 1 9 1988
E
D

88 9 19 240

Development of Performance Measures for the Navy Radioman

Herbert George Baker, Ph.D.
Navy Personnel Research and Development Center
San Diego, CA 92152-6800

Patrick Ford
Jack Doyle
Shelia Schultz
R. Gene Hoffman
Human Resources Research Organization
Alexandria, VA 22314-4499

Steven E. Lammlein, Ph.D.
Cynthia K. Owens-Kurtz
Personnel Decisions Research Institute
Minneapolis, MN 55414

Reviewed by
Pamela J. Kidder

Released and released by
John Pass, Ph.D.
Director, Personnel Systems Department

Navy Personnel Research and Development Center
San Diego, California 92152-6800

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS			
2a. SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.			
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE					
4 PERFORMING ORGANIZATION REPORT NUMBER(S) NPRDC TN 88-52		5. MONITORING ORGANIZATION REPORT NUMBER(S)			
6a. NAME OF PERFORMING ORGANIZATION Navy Personnel Research and Development Center	6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION			
6c. ADDRESS (City, State, and ZIP Code) San Diego, CA 92152-6800		7b. ADDRESS (City, State, and ZIP Code)			
8a. NAME OF FUNDING / SPONSORING ORGANIZATION	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER			
8c. ADDRESS (City, State, and ZIP Code)		10 SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO. 63707N	PROJECT NO Z1770	TASK NO. .001	WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)

Development of Performance Measures for the Navy Radioman

12 PERSONAL AUTHOR(S)
Baker, H. G., Ford, P., Doyle, J., Schultz, S., Hoffman, R. G., Lammlein, S., Owens-Kurtz, C.

13a. TYPE OF REPORT Interim	13b. TIME COVERED FROM 1982 TO 1986	14. DATE OF REPORT (Year, Month, Day) 1988 September	15. PAGE COUNT 29

16 SUPPLEMENTARY NOTATION

17	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Job performance measurement, joint-service research, hands-
05	09		on testing, job knowledge tests, rating scales

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This report details the development and tryout of performance measurement tests for the first-term incumbents of the radioman (RM) rating.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL Baker, H. G.	22b. TELEPHONE (Include Area Code) (619) 553-7639	22c. OFFICE SYMBOL Code 62

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. SECURITY CLASSIFICATION OF THIS PAGE
All other editions are obsolete.

UNCLASSIFIED

FOREWORD

The Navy Job Performance Measurement Program (subproject Z1770.001) constitutes a significant contribution to the Joint-Service Job Performance Measurement/Enlistment Standards Project. The Joint-Service Project has been mandated by Congress to link enlistment standards to job performance, which can be considered a landmark research thrust of the armed services. The present research has been funded primarily under PE 63707N (Manpower Control System Development) and project number Z1170 (Manpower and Personnel Development).

This report details the development and tryout of a comprehensive performance measurement system for the first-term incumbents of the radioman (RM) rating. Subsequent reports will address formal pilot testing and full-scale data collection.

Information contained in the report is intended to benefit the research and the operational RM communities. Ultimately, the outcome of the project will benefit the armed services, military and civilian research communities, and applied industrial/organizational psychology in general.

JOHN J. PASS
Director
Personnel Systems Department

v

SUMMARY

Problem

The armed services, in cooperation with the Department of Defense, are investigating the direct linkage of enlistment standards with on-the-job performance. This will require development and evaluation of performance measures and the comparison of performance on these measures with scores attained on selection and classification tests.

Objectives

The objectives of the work reported herein were to: (1) develop a comprehensive test package with which to assess the technical proficiency of first-term Navy radiomen (RM) personnel; and (2) tryout selected measures in the relevant test setting.

Approach

Measures were developed using job analysis, questionnaire data, subject matter expert (SME) workshops, and advisory panel review. A test package consisting of a hands-on test, a written job knowledge test, and a set of rating scales were administered to a small sample of first-term RMs.

Results

Try out results provided evidence that the performance measurement system (1) is well grounded in the RM real-world job, (2) is regarded as an acceptable job sample by RM SMEs, and (3) can be feasibly administered in appropriate settings.

Conclusions and Recommendations

The measures evaluated in this tryout are adequate for the assessment of first-term RM job performance. Development of the additional measures to be contained in the full test package should be completed. Prior to the full-scale validation study, it is recommended that the complete RM performance measurement system undergo a carefully controlled pilot test using a sufficient number of test subjects to evaluate the reliability and objectivity of the tests and to make necessary revisions.

CONTENTS

LIST OF FIGURES

INTRODUCTION

Problem

The armed services, in cooperation with the Department of Defense, are investigating the feasibility of directly linking enlistment standards with on-the-job performance. This will require development and evaluation of performance measures and the comparison of performance on these measures with scores attained on selection and classification tests.

Objectives

The objectives of this phase of the project were to (1) develop a comprehensive test package with which to assess the technical proficiency of first-term Navy radiomen (RM) personnel; and (2) tryout selected measures in the relevant test setting.

Background

Previous reports detail the research strategy and purposes of the Joint-Service Job Performance Measurement (JPM)/Enlistment Standards Project (Office of the Assistant Secretary of Defense, 1982), and the origin and scope of the Navy JPM Program (Laabs & Berry, 1987).

The objectives of the Joint-Service Project are to: (1) develop prototype methodologies for measuring job performance and (2) link enlistment standards to on-the-job performance. The Joint-Service research strategy currently focuses on measures of technical proficiency and the evaluation of surrogates for hands-on tests. The Navy JPM Program supports the Joint-Service project and incorporates three major efforts: (1) Performance-Based Personnel Classification (PBPC), which directly addresses the Joint-Service objectives; (2) development of an automated JPM data base; and (3) several technology-base research efforts aimed at resolution of important performance measurement issues.

In the PBPC effort, performance measures are being developed for seven ratings: (1) machinist's mate (MM), (2) radioman (RM), (3) electronics technician (ET), (4) operations specialist (OS), (5) fire controlman (FC), (6) electrician's mate (EM), and (7) gas turbine systems (GS). The aviation machinist's mate (AD) rating was added for a special demonstration of Joint-Service Project technology transfer potential. These eight ratings encompass over 25 percent of the Navy enlisted force, are among the 26 most critical Navy ratings, and use more than half of the 10 Armed Services Vocational Aptitude Battery (ASVAB) composites in the Navy's classification and assignment system (Laabs & Berry, 1987).

Development and administration of performance measures in each of the ratings constitute a substantial research project. Work in each rating will follow four major phases: (1) critical task selection, (2) test development/tryout, (3) pilot test, and (4) validation study. All four phases have been completed on the first rating (MM), and Phase 1 has been completed for the RM rating (Lammlein & Baker, 1987), with Phase 2 being the subject of this report.

1

CRITICAL TASK SELECTION

Lammlein and Baker (1987) described the identification of critical RM tasks and the selection of a subset on which the RM tests would be based. This process included a job analysis to compile a preliminary list of critical tasks performed by first-term RMs. "Critical" was defined as those tasks that are: (1) performed by a sizeable number of incumbents, (2) important to mission success, (3) characterized as having at least moderate variance in performance, (4) representative of the entire job domain to the greatest extent possible, and (5) subsumed in the first-term RM job across varied duty assignments. In addition, the tasks were restricted to those involving technical proficiency, which could be accomplished on an individual basis.

A set of 124 job tasks was identified as the technical domain of the first-term RM job. To isolate the more critical tasks in this set, a survey was conducted of first-term RMs and their supervisors (N = 500 each) in large and small shore establishments and on a wide variety of surface ships. The identified critical tasks were reviewed by the Quality Control Review Panel, a group of RM job experts serving the project in an advisory capacity. This panel ultimately selected 22 critical tasks that were incorporated in the test items based on several criteria: (1) criticality to the Navy's mission, (2) frequency of performance by first-termers, (3) variability in correct performance, and (4) adequacy and comprehensiveness of a job sample (i.e., that the collection of tasks was sufficient to stand as the core of first-term RM duties).

In some cases, selected tasks were integrated to make test conditions more realistic. For example, the panel selected three tasks that are rarely performed in isolation--set up crypto equipment, set up teletypes, and patch equipment pieces together. These tasks are typically performed in conjunction with each other when the RM establishes a system to transmit or receive. Based on the guidance of the Navy RM "A" School instructional staff, the three selected tasks were subsumed within two more global test items: establish a system to receive and establish a system to transmit.

The critical tasks form the basis for the development of the job performance test package for the RM rating. Ultimately, four types of measures are to be developed for the project. These include:

1. A hands-on test, or job sample test, which takes place in settings identical or highly similar to the job setting, wherein test subjects actually perform a job task.

2. Written tests of two varieties: (a) a task test, in which the items correspond task-by-task with the hands-on test; and (b) a job knowledge test (JKT), which addresses general knowledge required by the first-term RM.

When used for appropriate kinds of job tasks and linked firmly to knowledge-based task elements, JKT have wide applicability, acceptable validity, and are exceptionally efficient. A number of advantages accrue including: (a) economy (little equipment or personnel support is required; therefore, large groups of personnel can be tested in a relatively short period of time); (b) applicability (a knowledge test is actually a preferred method of testing for tasks that involve cognitive skills such as decision making, problem solving, and related applications of rules and principles); and (c) domain coverage (a greater breadth of task coverage can often be achieved with a knowledge test than with other methods because it is less subject to the environmental and safety conditions that often constrain hands-on tests).

3. A set of rating scales administered at the peer and supervisor levels that address two areas of the first-term RM job: (a) performance on the same critical tasks addressed by the hands-on and written tests; and (b) global job performance categories, which take into account not only technical proficiency but also non-technical performances that are required of all Navy enlisted personnel.

4. An overall performance rating (OPR), assigned by the test administrator upon completion of the hands-on testing, which is a subjective evaluation of the subject's overall performance. OPRs are not constrained to go, no-go gradings, but may include judgments concerning motivation, work attitudes, safety consciousness, and the end result (i.e., did the action taken result in an operational system?).

Figure 1 shows the composition of the performance measurement system.

- Hands-on Component Test
- Written Job Knowledge Tests
  - Written Task Test
  - General Knowledge Test
- Rating Forms (completed by peer and supervisor)
  - Task Rating Form
  - Global Rating Form
- Overall Performance Rating Form

Figure 1. List of measures.

APPROACH

A flowchart of the major steps in the development of the various measures, their tryout, and refinement is shown in Figure 2.

Development of the Hands-on Test

A hands-on test item was developed task-for-task with the final list of critical tasks approved by the Quality Control Review Panel (QCRP). Where items are equipment-bound, the radio gear was chosen that is found most frequently in use across ship types and shore installations.

Each item was composed of observable, behavioral steps toward its completion. The behavioral steps for proper execution of the items were specified in SME workshops by the job experts, guided by manuals, written procedures, Navy policies, and safety precautions. As all tasks are rather straightforward and pose minimal hazard to personnel or equipment, decision rules were limited to two: (1) each step must be faithful to policy, and (2) each step must be readily observable.

Where materials were required in a test item, sample messages, routing guides, and dummy classified materials were created. Specifically developed perforated tapes also ensured standardized test item content.
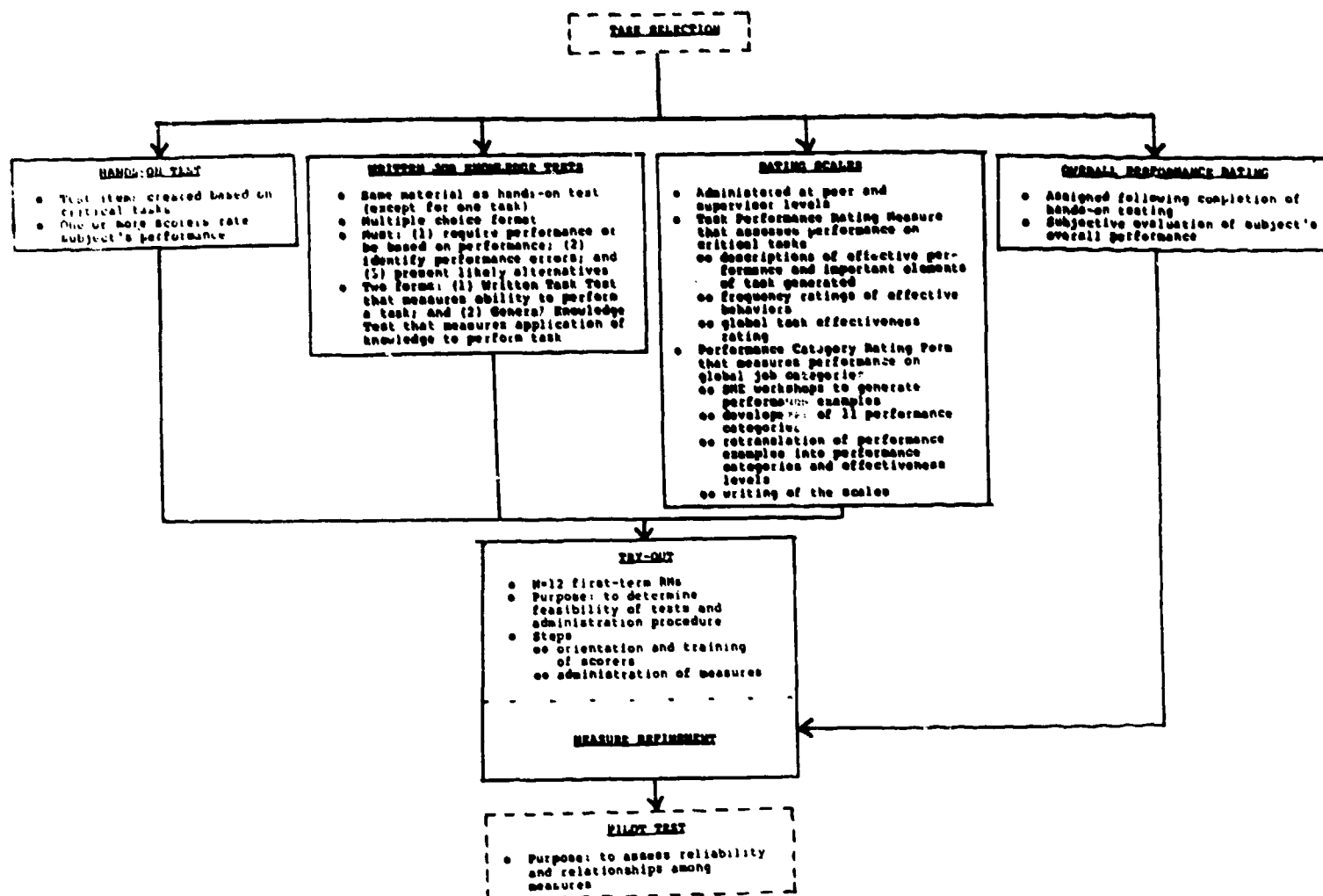
Figure 2. Development, tryout, and refinement of test package.

The array of hands-on test items is summarized in Table 1. This table shows which selected tasks have been integrated into a test item and describes the content of each hands-on test item.

Score sheets were developed for each test item. Score sheets contain a set of dichotomously scored elements corresponding to steps done correctly or incorrectly or to characteristics of task products that are acceptable or unacceptable. An example of a hands-on test item is found in Appendix A.

Development of the Written Job Knowledge Tests

Two types of written tests were developed, a task test and a general knowledge test. The difference between the types of written tests lies in the nature of the information that the items are attempting to elicit. The written task test attempts to measure ability to perform the task while the general knowledge test measures the application of knowledge to perform the task.

Table 1

Overview of Hands-on Tests

Selected Tasks	Test Item	Scope
Screen incoming messages for correct address, precedence, security classification, etc. Monitor channel number continuity for message traffic. Receive classified message traffic. Recognize and comply with special message handling procedures.	Act as broadcast operator	Twenty-one messages are transmitted to two teletypes. The messages represent the following range of characteristics: 5 general, 7 addressed and guarded, 4 addressed but not guarded, 5 classified, 2 cancelled, and 2 garbled or incomplete. The Radioman opens the circuit log for each channel, processes messages, and closes the circuit logs.
Maintain communications center message files.	Log messages File messages	The initial draft of the test had the Radioman prepare logs and files for the general and addressed messages received in Act as Broadcast Operator. After the pilot the test was divided: Log messages. The Radioman puts the messages in precedence order and logs addressed and general messages. File messages. The Radioman prepares fillers for general messages, files general messages in one file, and files addressed messages and general message fillers in another file.
Use routing guide to determine distribution or routing of incoming messages. Route classified messages (excluding CMS) to appropriate personnel. Manually route messages to appropriate destinations.	Manually route messages	Working with four messages from Act as Broadcast Operator and a routing guide, the Radioman determines who gets copies and how many copies are required.
Change paper/ribbons on teletypes and printers.	Change paper/ribbons on teletypes and printers	The Radioman takes out and replaces paper, printer ribbon, perforator ribbon and perforator tape.
Set up crypto equipment. Set up teletypes. Patch communication equipment pieces together.	Establish system--Golf Establish system--November	The Golf system is the most widely used system to transmit; the November system is the most widely used to receive. The Radioman reads the circuit board to determine what equipment is available and what frequency to use then selects the equipment, connects the components (using patch cords or dials as appropriate), and adjusts settings. Both tests have a 20 minute time limit to establish the system.
Inventory confidential materials (excluding CMS). Inventory secret materials (excluding CMS). Destroy secret materials (excluding CMS).	Inventory Confidential/Secret Documents	The Radioman conducts an inspection of publications including filling out an inventory form. The Radioman then describes the procedure to destroy a secret document by burning.
Perform preventive maintenance on receiver.	Perform preventive maintenance on receiver	The Radioman follows a Maintenance Requirement Card (MRC) for the receiver (R-1051/D). The test includes safety procedures for checking and cleaning electrical connections, checking mechanical components, and cleaning the receiver. The 12 minute time limit is tight.
Perform preventive maintenance on transmitter.	Perform preventive maintenance on transmitter	As with the receiver, the Radioman follows an MRC for the transmitter (AN/URT-23). The test covers inspecting the air filters and testing the alarm circuit. The 12 minute time limit is comfortable.
Proofread outgoing messages prior to transmission Select/use relevant general publication instructions and directions.	Prepare message--DD173	The Radioman types a joint message form (JMF). Scorers check the completed, corrected form.
Prioritize outgoing message according to precedence and time of receipt.	Prioritize outgoing messages.	The Radioman first sorts seven messages on a JMF in order of priority for transmission. The messages contain a range of precedence for addressee and time of receipt as well as irrelevant ranges of precedence for information copy and classification. The Radioman also states the time objective for each precedence level.
Type/format/edit messages on teletype.	Type/format/edit message	The Radioman writes the header and cuts a tape for a given message for transmission by teletype. Scorers check the written header and the tape for format and accuracy.
Verify outgoing messages on DD173 for completeness, accuracy, format, and releasing signature.	Verify outgoing messages	The Radioman is shown six messages on the JMF. Each message has at least one format error. The Radioman identifies the errors.

5

Written Task Test

All but one hands-on task was covered in the written task test. The exception was: Change paper/ribbons on teletypes and printers. This task was not included because it is an exclusively psychomotor task that does not lend itself to a written test format. Other than this exception, the written task test covers the same material as the hands-on test, only in a multiple-choice format. The multiple choice format was chosen for the written test because it is familiar to most sailors and is relatively easy to score. Since the test seeks to measure an RMs ability to perform a task, each item is intended to exhibit three characteristics:

> Require performance or be based on performance. In some cases the only difference between a hands-on test and a written test is that hands-on items call for recall responses and written tests call for recognition responses. For example, the written test for Manually Route Messages requires the RM to read a routing guide and determine the number of copies. That is the same performance required by the hands-on test (and the job itself). In most cases, though, the item is a question on an aspect of how the task is done. Even these items are intended to be based on task performance.

> Identify performance errors. To ensure that items are based on task performance, each item is intended to relate directly to some step or set of steps in the task. The focus for an item was determined by identifying causes of error. Four causes of error were considered: the RM did not know where to perform (usually location of components), when to perform (usually sequence), what the product of the correct performance is (usually proofreading), or how to perform (technique).

> Present likely alternatives. In the same way that items were to focus on performance errors, incorrect alternatives were intended to reflect errors that are possible and that do occur. In addition, the incorrect alternatives were to be wrong rather than less desirable than the correct alternative.

General Knowledge Test

The RM general knowledge test was based upon the same information as that used in hands-on and written task tests. The following general guidelines were observed in the development of the general knowledge test:

1. An attempt was made to restrict the stem of the item to two lines. Throughout, there was an effort to minimize the reading skills necessary to take the test.

2. The stem of each item was designed so that the item could be answered based on the stem alone; that is, without reference to the alternatives.

3. Tests were checked for inter-item cueing.

4. Each correct alternative was verified as correct by a citable reference.

6

Inherent cueing of multiple choice options, particularly between items, often makes it difficult to develop likely and plausible yet clearly wrong alternatives. Incorrect alternatives were limited to four by format design, but in some instances only two or three "real world" alternatives were possible and these were all that were listed. Therefore, three to five choices were developed for each item, with the answer being a single correct response. An example of the written task test is found in Appendix B.

Development of the Rating Scales

Rating scales are to be developed for the peer and supervisor levels. The two rating formats are: performance on critical tasks, and performance on global job performance categories. These require different developmental steps.

Behavior Summary Scales

The development of the behavior summary scales involved an additional job analysis through the collection of performance examples (sometimes termed "critical incidents"). These performance examples are illustrations of first-term RM performance at different levels of effectiveness, and were used to inductively generate the performance categories as well as to anchor the behavior summary scales with concrete descriptions of performance at different levels. The behavior summary scale format was chosen because of its inductive developmental process that is readily understandable and acceptable to job experts.

The development of the behavior summary scales proceeded in four steps (Bernardin & Smith, 1981; Smith & Kendall, 1963): (1) generation of performance examples, (2) development of performance categories, (3) retranslation of performance examples into performance categories and effectiveness levels, and (4) writing of the scales.

Generation of Performance Examples. Six SME workshops were held (in Norfolk, VA; San Diego, CA; and Stockton, CA). In these workshops, SMEs (RM supervisors in at least their second term) were asked to provide performance examples. Of the 95 SMEs who participated, 30 percent were stationed aboard ship and 70 percent were stationed on shore. Virtually all had worked in both types of assignments.

Training was provided on writing performance examples. For each example, participants were asked to provide descriptions of the circumstances in which the behavior occurred, the actual behavior observed, and relevant consequences that followed from the behavior. In addition, they were asked to rate the effectiveness level of each example and assign it to a job performance category. As participants wrote performance examples, the examples were collected and feedback was provided by research staff.

A total of 1,290 performance examples were generated in these workshops. These were then edited, corrected for grammatical errors, unnecessary detail, conciseness, and elimination of non-behavioral examples.

Development of Performance Categories. The development of a set of comprehensive, relatively homogeneous performance categories was an on-going process that began during the performance example workshops. Performance examples gathered from two workshops were sorted according to content similarity. These initial categories were reviewed by SMEs at the next workshop who considered criteria such as comprehensiveness, clarity, and applicability to different duty assignments (e.g., ship and shore). Their suggestions (and the examples they provided in their workshop) were

incorporated in a new set of categories, which was then reviewed in a subsequent workshop, and so on. Through this iterative process, a final set of categories was converged upon. Eleven performance categories, with associated definitions, resulted. They are shown in Figure 3.

Equipment and System Operations

Operating and caring for equipment properly; being alert to equipment problems and taking appropriate response to them; performing operating maintenance such as reloading paper, ribbon, and tape, clearing jams, etc.; patching individual pieces of equipment together to form a system; preparing antennas for use; tuning transmitters and receivers; setting up crypto equipment and changing key material; loading computer tapes; performing start-up sequences; performing system checks; troubleshooting system problems.

Circuit Communications

Establishing and maintaining circuit communications; making circuit responses according to proper procedures; monitoring circuit to ensure channel number continuity, message quality, etc. and taking appropriate corrective action as required; performing channel checks and other measures to ensure circuit reliability; finding new frequencies as required; troubleshooting circuit problems.

Processing Messages

Screening messages for precedence, classification, special handling requirements, etc. and taking appropriate action as required; routing, coping, and distribution incoming messages; preparing and transmitting outgoing messages; typing messages according to proper formats; proofreading/verifying messages for accuracy, format, completeness, releasing authority, etc.; resolving discrepancies when messages are rejected by computer; cutting tapes; servicing messages.

Filing, Record-keeping, and Clerical Duties

Filing and retrieving messages, reports, etc.; maintaining files; purging files and destroying materials as appropriate; maintaining and updating logs and status boards; preparing lists, reports, forms, etc.; conducting inventories and page checks of publications, instructions, bulletins, etc. and resolving discrepancies; making changes and corrections to publications and maintaining associated lists; maintaining supplies; processing records and funds for commercial traffic; performing miscellaneous typing such as reports, non-message forms, etc.; receiving telephone calls.

Equipment Maintenance and Repair

Following proper preventive maintenance procedures; tagging out equipment properly; performing equipment inspections; being alert during routine maintenance to additional equipment problems requiring attention; using correct tools, parts, lubricants, solvents, etc.; performing maintenance according to schedule; repairing equipment when required.

Figure 3. Job performance categories.

8

Security Mindedness

Protecting the security of classified material against compromise; ensuring proper handling, storage, and destruction of classified material; using secure communications procedures; distributing classified material only to appropriate persons; protecting security during delivery runs, visits by outside personnel, burn runs, etc; limiting access to classified spaces.

Safety Mindedness

Adhering to safety procedures and taking appropriate safety precautions; conducting thorough safety inspections; being alert to safety violations and hazards and taking appropriate action in response to them; being skilled in first-aid and other emergency procedures and applying them as necessary.

Acquiring and Using Technical Knowledge/Keeping Up-to-date

Staying knowledgeable and skilled in job responsibilities; seeking and using publications, SOP, instructions, manuals, etc., to perform job; seeking job information from others; pursuing opportunities to expand job knowledge and skills; qualifying for new positions/responsibilities.

Working With Others

Working with co-workers, supervisors, subscribers, etc. in a constructive, harmonious manner; helping out others on the job as appropriate; keeping others informed of relevant job information; preparing training and training others; monitoring progress of trainees and ensuring that training requirements are met; delegating duties as appropriate; supervising others.

Maintaining Living/Work Areas

Keeping work and living areas orderly and clean; cleaning floors, decks, etc.; performing field day cleanup tasks; securing loose objects on ship.

Conscientiousness, Extra Effort, and Devotion to Duty

Reporting on time and fit for duty; putting in effort to get jobs/assignments done; volunteering for duties; working long hours or in unpleasant conditions as necessary; behaving in a controlled, professional manner; representing appropriate appearance and uniform; respecting authority of chain-of-command.

Figure 3. (Continued).

9

Retranslation. "Retranslation" is essentially the process of verifying performance category assignments and determining the effectiveness levels of performance examples. It is accomplished by having SMEs assign each performance example to a category and rate its effectiveness level. Agreement across SMEs is then assessed, and those examples for which there is substantial agreement are used to typify the respective effectiveness levels of each category.

Retranslation was accomplished in a series of workshops held in Norfolk, VA; San Diego, CA; and Stockton, CA. A total of 70 first-term RM supervisors, second-term and above, participated. Each SME was provided with the performance categories and definitions and a retranslation booklet containing one-third of the performance examples. The order of the booklets was varied so that approximately equal numbers of SMEs completed each booklet and a number of SMEs started with the second half of the booklet). SMEs were first instructed to read through the performance categories and definitions in order to become familiar with them. They were then instructed to read each performance example and determine which category is represented by the behavior described. The letter of the category was to be written next to the example. The effectiveness level of the example was then to be rated, ranging from 1 (low) to 9 (high). SMEs then filled out the retranslation booklet they were given at the start of the workshop.

Analysis of retranslation data included the number of respondents who retranslated each example, the percentage of respondents who assigned each example to each performance category, and the mean and standard deviation of the effectiveness level rating for each example.

Rating Scale Development. The retranslation results were used to select performance examples for each performance category, based on the following primary criteria: (1) substantial agreement as to category assignment (60% or more of the respondents assigned the example to that performance category), and (2) effectiveness level (the standard deviation of the effectiveness level ratings was less than 1.5). In addition, the attempt was to cover the entire range of effectiveness for each category with selected examples. For some categories there were few or no examples from the middle level of effectiveness that fulfilled the second criterion. In such cases, the acceptable standard deviation was raised to 2.00, and additional information was interpolated from low- and high-effectiveness examples in those categories.

Based on the content of the performance examples, summary statements were written for high, medium, and low effectiveness in each category. Emphasis was placed on ensuring that all important behaviors brought out in the examples for each category were included. An additional concern was to unambiguously distinguish between the levels of effectiveness.

In addition, specific performance examples illustrating each level of effectiveness for each category were chosen to complement the summary statements. These examples were chosen for (1) behavioral representativeness to the broader set of examples, (2) high agreement on category and effectiveness level, and (3) conciseness.

Slight modifications were made to the category definitions based on the retranslation results. In some cases, examples dealing with a certain area of a category definition did not retranslate well into that category, and the definition was revised accordingly.

10

Thus, the behavior summary scale for each category consisted of a title and definition, and, for each of three effectiveness levels, a summary statement and illustrative performance examples. A seven-point rating scale was used.

The resulting Performance Category Rating Form contained three sections. The first section requested background information such as the rater's and ratee's pay grade, their working relationship (i.e., peer, first-line supervisor, or second-line supervisor), the length of time they have worked together, and a judgment of the rater's confidence in his/her ratings. The second section contained the behavior summary scales for the 11 categories. Finally, a global performance rating was requested in the third section. Appendix C contains an example of the rating scales developed in this research effort.

Task Performance Rating Measure

The development of the task rating scales primarily involved SME workshops to isolate the critical performance behaviors associated with effective performance on each of the critical tasks.

In contrast to the Performance Category Rating Form, which assesses performance on general categories of first-term RM performance that include technical, non-technical, and Navy-wide areas, this rating instrument specifically addressed performance on the critical tasks selected in the job analysis, a much narrower focus.

Description of the Format. A different rating form was used for rating the critical tasks. The reason for this is that behavior summary scales are better adapted to more general areas of performance. Furthermore, to have developed the task scales through the same process as the behavior summary scales would have required collecting a prohibitive number of performance examples.

The scale format chosen essentially incorporates elements of the behavior observation scaling (BOS) format (Hough, Sevy, & Dunnette, 1983; Lammlein & Hough, 1984; Latham & Wexley, 1982). For each task or other rating unit, descriptions of effective performance on important elements of the task or rating unit are required. These are referred to as "task rating items." These rating items are each rated for the frequency with which the ratee displays the effective behavior; thus the similarity to the BOS format.

Following these frequency ratings for all the rating items of a task, a global task effectiveness rating is also made. This is a relative rating for which the ratee is compared to all other performing the task. If the ratee is among the most effective at performing the task, a high rating is assigned. Similarly, if the ratee is poorer than most at performing the task, a relatively low rating is justified.

Development. During each performance example generation and retranslation SME workshop, 4-6 of the SMEs with varying duty backgrounds were asked to work on the critical task rating form. This was done only after they had written some performance examples so that their focus would be on behaviors associated with effective performance.

The first SME group generated the task rating items for each critical task; that is, the specific behaviors associated with effective performance on all important components of each task. Subsequent groups then reviewed these items and made changes, additions, or deletions as necessary.

11

In some cases, SMEs recommended wording changes to the tasks to clarify what would be rated. For example, "Set up crypto equipment" was changed to "Perform initial set-up of crypto equipment."

Following the development of the task rating items, the rating form was constructed. A frequency scale was developed for rating the task items, with items ranging from 1 ("never or rarely") to 6 ("always"). The overall effectiveness scale ranged from 1 ("least effective--approximately 1-10%") to 5 ("most effective--approximately 91-100%"). Both scales had an option for "not part of job or cannot rate."

To maintain comparability across the RM performance measures, tasks were grouped on the rating form according to the job areas developed for the hands-on and knowledge tests (e.g., broadcast operator, establish system, etc.). An overall job area effectiveness rating was included for each, using the same rating scale as the overall task effectiveness ratings.

Development of the Overall Performance Rating (OPR)

The OPR was based directly on the hands-on test. Item for item, the OPR required the test administrator to rate the test subject's global performance on the task. OPR ratings are based on a 1 (far below the acceptable level of proficiency) to 5 (far exceeded the acceptable level of proficiency) point scale. This instrument was not developed for the tryout and was not included as part of the test package. An OPR rating sheet is contained in Appendix D.

Tryout

The tryout was conducted in San Diego, California in September of 1986. Each Radioman took the general knowledge test and the hands-on test. The written task tests and the overall performance rating were not completed by the time of the tryout, and thus were not used. Also, the peer rating forms were not used. RM supervisors rated the test subjects' everyday performance using both the Performance Category Rating Form and the task rating materials. The primary purpose of the tryout was to determine the feasibility of the tests. The tryout was more an evaluation of feasibility and administration procedures than of the rating measures.

Subjects

Twelve first-term RMs participated in a tryout. All had graduated from Radioman "A" School and were in ship assignments. Nine were in garde E-3, three were in grade E-4. Eleven senior petty officers, supervisors of the RMs, rated the 12, and two ratings were obtained for most of the 1...

Procedures

Hands-on Tests. The hands-on tests were administered in the Team Trainer at the Naval Training Center, San Diego. This is a simulated radio shack operationally parallel to a radio shack on a ship. All components of the radio shack are operating production models of equipment.

The scorers for the tryout were two staff members who had been hired specifically for the testing. Both had recent experience as instructors at the RM "A" School. Since they were both experts on the tasks, scorer training focused on test procedures. An

orientation emphasized the need for complete information on each RM's performance down to the step level and the importance of maintaining a consistent, neutral attitude toward each radioman tested.

The scorers were made familiar with the specific measures by having one scorer perform the task while the other completed the scoresheet. Besides familiarizing the scorers, this task resulted in minor modifications to each task test. The typical modification was to delete measures for which failure to perform was not a clear violation of correct procedure.

The scorers administered the modified tests to the 12 radiomen. Each radioman performed each task while being evaluated by one scorer. A scorer was paired with a radioman and administered all the tests that radioman could complete in a 4-hour period. Because of time constraints, some radiomen did not complete all task tests. All subjects completed at least 10 tests; seven completed all 13.

Knowledge Test. The written test was administered by contractor personnel to all subjects in a classroom setting.

Rating Scales. Ratings were collected in workshops where the supervisors were briefed on the project and ensured confidentiality for their ratings (they were informed that the ratings would be used for research purposes only). The order of administration for the two rating measures was counterbalanced. The supervisors read to themselves the instructions for the first rating instrument, and then questions were answered. Following this, rater error training was provided. This consisted of describing halo error, recency error, stereotype error, and same-level-of-effectiveness error as a caution against committing them. Subsequently, the ratings were completed. A similar procedure was followed for the second rating instrument, except that the error training was abbreviated.

Supervisors were asked to comment on the rating instruments, especially for information regarding the clarity of instructions and scales, job coverage, and feasibility of providing such ratings.

Test Package Refinement and Further Development

The test package was refined through: (1) modifications to the instruments based on information gathered during the tryout; and (2) completing the development of the overall performance rating form and the written task test, and generating the peer-level set of rating scales.

RESULTS

Major elements of the RM job performance measurement system were tried out in appropriate settings. Measures (and their associated instructions) were revised accordingly.

Hands-on Test

The tryout demonstrated that the tests were feasible although they required substantial resources per radioman tested. The set of tests required an average of 3 hours to perform, not including time for instructions, equipment set up, or equipment maintenance. The ratio of one hands-on scorer per two RMs per 8-hour day was just barely

acceptable at the start of testing. Fortunately, efficiency increased after the first day so a higher proportion of RMs completed all the task tests. Thus, no tasks were recommended for deletion based on feasibility considerations.

A question related to feasibility concerned whether the test requirements were acceptable to both the RMs and the scorers as a sample of the first-term RM job. The equipment, conditions, and instructions conduced to the behaviors the test was designed to address and none of the participants voiced major objections to the test content or organization. The Log Incoming Messages and File Incoming Messages will be combined into one test item, Log/File Messages. Likewise, Establish System-Golf and Establish System-November will be combined into one test item, Establish System (Golf & November). These changes will result in a total of 12 items.

The major purpose of the tryout was to assess timing, administrative feasibility, etc. Obviously, the very small sample precluded statistical analyses of any sort.

General Knowledge Test

The 106-item general knowledge test was administered in order to review the proportion of correct responses (i.e. difficulty parameter), as well as each item's relevance to the test and the quality of each item. Test scores ranged from a high of 67 to a low of 39. Mean performance was moderate, with an overall mean of 54; four of the test results fell below the mean. No changes to the items were deemed warranted.

Rating Scales

Some modifications were made to the Performance Category Rating Form and the Task Rating Form following the tryout. The comments on the behavior summary scales were, in general, quite positive. Two suggestions were subsequently incorporated into the scales. "Not part of Job or Cannot Rate" was added as an option to the 1-7 effectiveness rating because some supervisors believed they could not rate some RMs on every category. In addition, the label on the middle level of effectiveness was changed from "Fully Adequate" to "Average Effectiveness." The reason for this change was that the supervisors regarded fully adequate as, in fact, exemplary performance and thus were confused by two separate effectiveness categories that both implied the highest level of performance.

Data analyses were not conducted on the rating data due to the small sample size involved. Of primary interest, rater, were the comments on the rating instruments.

The reactions to the task rating form were quite positive, resulting in only two changes. One involved adding a general preventive maintenance task (not equipment specific) to supplement the two critical tasks involving preventive maintenance in specific pieces of equipment. SMEs thought this to be necessary because some first-term RMs may perform preventive maintenance, but not on the two pieces of equipment (receivers and transmitter) mentioned in the rating form.

The other change consisted of making the frequency rating a five-point scale. The reason for this is that some people assigned "6" ratings for overall effectiveness, suggesting that using two rating scales with different numbers of rating points was confusing.

Tryout of the measures includes assessing operational community response. In most cases, SMEs enjoyed participating in the process of defining effective, ineffective, and average performance. The supervisors who made ratings gave very positive feedback on the rating measures, especially with regard to clarity, job relevance, comprehensiveness, and appropriateness for first-term RMs.

Completion of the Performance Measurement System

Following the same procedures used to develop the written test and ratings included in the tryout, development was completed on the Written Task Tests, the peer-level rating scales, and the Overall Performance Rating Form. These were incorporated into the RM comprehensive performance measurement system that will undergo pilot testing in the next phase of research.

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this tryout was to make preliminary evaluations of three different types of performance measures. Conclusions to be drawn from the tryout results are necessarily tentative due to the small sample size involved. However, it is apparent that the performance measurement system is well grounded in the RM-real-world job because it is accepted as by RM SMEs as a sample of the first-term RM's job.

Prior to the full-scale validation study, it is recommended that the complete RM performance measurement system undergo a carefully controlled pilot test using a sufficient number of test subjects to evaluate the reliability and objectivity of the tests and to make necessary revisions.

15

REFERENCES

Bernardin, H. J., & Smith, P. C. (1981). A clarification of some issues regarding the development and use of behaviorally anchored rating scales. Journal of Applied Psychology, 66, 458-463.

Laabs, G. J., & Berry, V. M. (1987, August). The Navy Job Performance Measurement Program: Background, inception, and current status (NPRDC Tech. Rep. 87-34). San Diego: Navy Personnel Research and Development Center.

Lammlein, S. E., & Baker, H. G. (1987, February). Developing performance measures for the Navy radioman (RM): Selecting critical tasks (NPRDC Tech. Rep. 87-13). San Diego: Navy Personnel Research and Development Center.

Lammlein, S. E., & Hough, L. M. (1984). Content validation study for Library of Congress Special Policy (Institute Rept. No. 92). Minneapolis: Personnel Decisions Research Institute.

Latham, G. P., & Wexley, K. N. (1982). Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley.

Hough, L. M., Sevy, B. A., & Dunnette, M. D. (1983). Development and criterion-related validation of a Library of Congress Special Policy selection system (Institute Rept. No. 89). Minneapolis: Personnel Decisions Research Institute.

Office of the Assistant Secretary of Defense (MRA&L). (1982, December). Joint-Service efforts to link enlistment standards and job performance: First annual report to the House Committee on Appropriations. Washington, DC: Author.

Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

APPENDIX A

EXAMPLE OF HANDS-ON TEST ITEM (TASK)

Scorer: _____   Radioman: _____

Date: _____   ID #: _____

BROADCAST OPERATOR
(Screen Incoming Messages, Monitor Channel, Receive Message Traffic)

INSTRUCTIONS TO RADIOMAN: For this test you act as a broadcast operator and monitor your assigned channels for messages. You must open your circuit logs, receive message traffic, process the messages and maintain the Broadcast File. Are you ready to receive messages? Begin.

PERFORMANCE MEASURES:	GO	NO-GO

Opened Circuit Log (HMAA)

1. Recorded current RADAY on Broadcast Circuit Number Log.   ___ ___

2. Recorded broadcast channel designator on Broadcast Circuit Number Log.   ___ ___

3. Recorded channel designator as HMAA.   ___ ___

4. Drew line above first broadcast channel number to be copies.   ___ ___

5. Drew diagonal lines through previous numbers not copied (X).   ___ ___

6. Recorded first 3 digits of channel sequence number indicated on first broadcast number copied.   ___ ___

Opened Circuit Log (HMCC)

7. Recorded current RADAY on Broadcast Circuit Number Log.   ___ ___

8. Recorded broadcast channel designator on Broadcast Circuit Number Log.   ___ ___

9. Recorded channel designator as HMCC.   ___ ___

10. Drew line above first broadcast channel number to be copied.   ___ ___

11. Drew diagonal lines through previous numbers not copied (X).   ___ ___

12. Recorded first 3 digits of channel sequence number indicated on first broadcast number copied.   ___ ___

APPENDIX B

EXAMPLE OF WRITTEN TASK TESTS

Example of Written Task Test

Broadcast Operator

1. Which best defines a separator?

   A. Is part of the heading component.

   B. Separates the text from other parts of the message.

   C. Is part of the ending procedure.

   D. Is part of the text.

2. Which of the following does the four-letter broadcast channel designator HMAA identify?

   A. EASTPAC, submarine, channel 1.

   B. WESTPAC, fleet multichannel, channel 3.

   C. EASTPAC, fleet multichannel, channel 1.

   D. Lant and Med, general CW, channel 3.

3. How often should the broadcast circuit number log be closed out?

   A. Daily @ 2359Z hrs.

   B. Monthly @ 2359Z hrs, 30th day of month.

   C. Yearly @ 2359Z hrs, last day of calendar year.

   D. Daily @ 0001Z hrs.

4. Which is the correct way to identify an unclassified message not addressed to your command when filling out your broadcast circuit log?

   A. 11 UECST

   B. 11 UECST

   C. 11 UECST

   D. (11) (U)ECST

5. Which of the following properly identifies a cancelled transmission on your broadcast circuit log?

   A. (11) UECST

   B. 11 BUST UECST

   C. 11 UECST ZES-2

   D. (11) (U)ECST

RADIOMAN ITEMS

1. The process of minimizing is the _____.

    a. proper use of correction tape

    b. preparation of a readdressal

    c. reduction and control of electrical message and telephone traffic during an emergency or exercise

    d. the precedence assigned to all types of message traffic which is not of sufficient urgency to require a higher precedence

2. Identify the following narrative message. (See diagram below)

JOINT MESSAGE FORM					SECURITY CLASSIFICATION UNCLASSIFIED				
PAGE 01 of 01	DTG RELEASER TIME		PRECEDENCE		CLASS	SPECAT	LMF	GR	ORIG/MSG IDENT
	DATE/TIME	MONTH	YR	ACT	INFO					
	201242Z	APR	83	RR		UUUU				1401242
BOOK		MESSAGE HANDLING INSTRUCTIONS							

FROM: COMNAVTELCOM WASHINGTON DC
    TO: ALCOM

UNCLAS   //N02319//
ALCOM 012/83

    a. Single Address

    b. Multiple Address

    c. General

    d. Book

3. The first line of a casualty report is identified by the letters _____.

    a. NSGIB/CAMREP/

    b. RSGID/CASRED/

    c. MSBIP/CALREP/

    d. MSGID/CASREM/

    e. MSGID/CASREP/

APPENDIX C

EXAMPLE OF RATING SCALES

Example of Task Performance Rating Form

FREQUENCY	OVERALL EFFECTIVENESS
N = NOT PART OF JOB OR CANNOT RATE 1 = NEVER OR RARELY 2 = OCCASIONALLY 3 = RATHER FREQUENTLY 4 = ALMOST ALWAYS 5 = ALWAYS	N = NOT PART OF JOB OR CANNOT RATE 1 = LEAST EFFECTIVE (APPROX. 1-10%) 2 = LESS EFFECTIVE (APPROX. 11-33%) 3 = EFFECTIVE (APPROX. 34-66%) 4 = MORE EFFECTIVE (APPROX. 67-90%) 5 = MOST EFFECTIVE (APPROX. 91-100%)

JOB AREA D: PREVENTIVE MAINTENANCE

Perform Preventive Maintenance on Receivers (Using MRCs)

1. Selects proper and current MRCs when performing preventive maintenance on receivers. ☐

2. Carefully and completely follows step-by-step maintenance instructions on MRCs when performing preventive maintenance on receivers. ☐

3. Observes MRC safety and tag-out precautions when performing preventive maintenance on receivers. ☐

4. Takes appropriate follow-up action (e.g., notifies supervisor, completes paperwork, etc.) to record completion of maintenance or discrepancies found when performing preventive maintenance on receivers. ☐

5. Performs preventive maintenance on receivers according to schedule. ☐

6. OVERALL EFFECTIVENESS at performing preventive maintenance on receivers (using MRCs). ☐

Perform Preventive Maintenance on Transmitters (Using MRCs)

1. Selects proper and current MRCs when performing preventive maintenance on transmitters. ☐

2. Carefully and completely follows step-by-step maintenance instructions on MRCs when performing preventive maintenance on transmitters. ☐

3. Observes MRC safety and tag-out precautions when performing preventive maintenance on transmitters. ☐

4. Takes appropriate follow-up action (e.g., notifies supervisor, completes paperwork, etc.) to record completion of maintenance or discrepancies found when performing preventive maintenance on transmitters. ☐

5. Performs preventive maintenance on transmitters according to schedule. ☐

6. OVERALL EFFECTIVENESS at performing preventive maintenance on transmitters (using MRCs). ☐

Perform Preventive Maintenance

1. Selects proper and current MRCs when performing preventive maintenance. ☐

2. Carefully and completely follows step-by-step maintenance instructions on MRCs when performing preventive maintenance. ☐

3. Observes MRC safety and tag-out precautions when performing preventive maintenance. ☐

4. Takes appropriate follow-up action (e.g., notifies supervisor, completes paperwork, etc.) to record completion of maintenance or discrepancies found when performing preventive maintenance. ☐

5. Performs preventive maintenance according to schedule. ☐

6. OVERALL EFFECTIVENESS at performing preventive maintenance (using MRCs). ☐

OVERALL EFFECTIVENESS AT PREVENTIVE MAINTENANCE. ☐

E. EQUIPMENT MAINTENANCE AND REPAIR

Following proper preventive maintenance procedures; tagging out equipment properly; performing equipment inspections; being alert during routine maintenance to additional equipment problems requiring attention; using correct tools, parts, lubricants, solvents, etc.; performing maintenance according to schedule; repairing equipment when required.

1 LOW	2	3	4 AVERAGE	5	6	7 HIGH

Fails to follow established preventive maintenance procedures; fails to perform routine checks and maintenance on schedule; improperly tags out equipment; incorrectly repairs and reassembles equipment; damages equipment through the use of incorrect tools or materials, or through improper handling.

Follows established procedures and uses appropriate tools and materials when performing preventive maintenance; completes preventive maintenance in acceptable time; performs equipment checks and inspections at scheduled intervals; tags out equipment properly; repairs and assembles equipment with acceptable speed and accuracy.

Performs preventive maintenance thoroughly, quickly, and on or ahead of schedule; is alert to additional equipment problems when performing routine maintenance and responds appropriately; correctly repairs and reassembles equipment in minimal time; notes and corrects irregularities in PMS cards; initiates procurement of needed parts.

EXAMPLES ILLUSTRATING LOW EFFECTIVENESS

- While performing preventive maintenance on transceivers, this RM used the wrong type of lube oil. The transceivers would not function properly as a result.

- This RM incorrectly performed minor repairs to the base and gears of deck edge antennas. The gear wheel froze in place when the antennas were lowered for testing.

EXAMPLES ILLUSTRATING AVERAGE EFFECTIVENESS

- Assigned to do preventive maintenance, this RM did the job completely, in order, and in acceptable time.

- This RM properly performed scheduled equipment inspections, noting any problems or discrepancies.

EXAMPLES ILLUSTRATING HIGH EFFECTIVENESS

- While cleaning an air filter during routine maintenance, this RM noticed some black marks on the inside of a transmitter. The RM reported this to maintenance personnel, who found and replaced some electronic parts that were arcing.

- After high winds broke a wire on an antenna, this RM took down the wire's aft end, repaired it, and reassembled the wire onto the antenna in minimal time.

How effective is each RM in this category?

1	2	3	4	5	6	7	N
1	2	3	4	5	6	7	N
1	2	3	4	5	6	7	N
1	2	3	4	5	6	7	N
1	2	3	4	5	6	7	N

APPENDIX D

OVERALL PERFORMANCE RATING SHEET

OVERALL PERFORMANCE RATING SHEET

DIRECTIONS FOR OVERALL TEST
ITEM PERFORMANCE RATING

INCUMBENT'S NAME
_____

RATER'S NAME
_____

Upon completion of each test item, please evaluate the incumbent's
overall performance for that particular item by using the following
scale. That is, enter the appropriate rating category (1 through 5) ·
in the space provided to the left of each test item.

---

5 Far exceeded the acceptable level of proficiency
4 Somewhat exceeded the acceptable level of proficiency
3 Met the acceptable level of proficiency
2 Somewhat below the acceptable level of proficiency
1 Far below the acceptable level of proficiency

PERFORMANCE CATEGORY	TEST ITEM
_____	Act as Broadcast Operator
_____	Log Messages
_____	File Messages
_____	Manually Route Messages
_____	Change Paper/Ribbons on Teletypes and Printers
_____	Establish System - Golf
_____	Establish System - November
_____	Inventory Confidential/Secret Documents
_____	Perform Preventive Maintenance on Receiver
_____	Perform Preventive Maintenance on Transmitter
_____	Prepare Message - DD173
_____	Prioritize Outgoing Messages
_____	Type/Format/Edit Messages
_____	Verify Outgoing Messages

DISTRIBUTION LIST

Chief of Naval Operations (OP-01B2)
Director, Accession Policy OASD (FM&P) (2)
Defense Technical Information Center (DTIC) (2)